Linear Regression with Movie Statistics

Andrew Auyeung

Outline

Methodology

What do the numbers say about a good movie?

Conclusion / Next Steps

Does Money Make Money?

Methodology

Data: BoxOfficeMojo

Sample: Domestic Movies from
2000-2019 with a large release
(1500+ theaters)

Target: Predicting Domestic Gross

Features

Strong Indicators	Weak Indicators
Historical statistics of Production Team	Proximity to Holiday Weekends
Release Date	Genre
Runtime	MPAA Rating

Feature Relationships

Low Budget, Low Domestic Gross

High Budget, Higher Domestic Gross
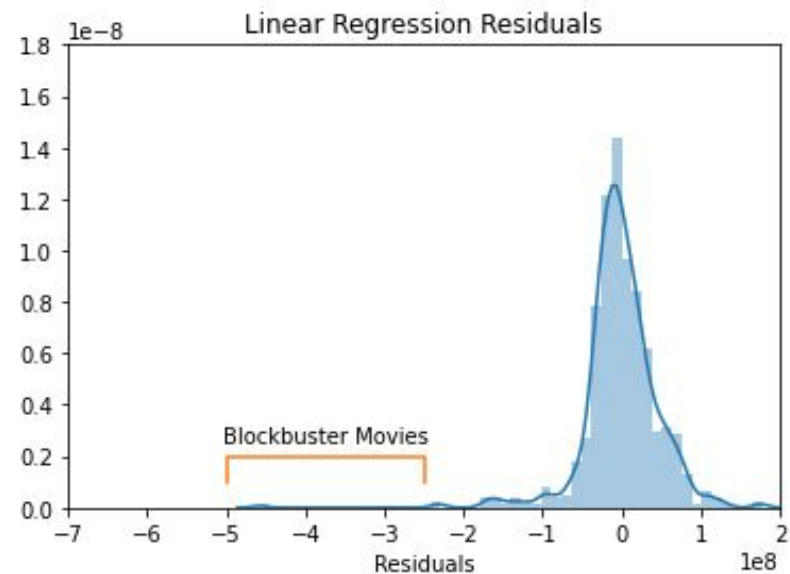
Feature Relationships

Big Ticket Actor Success
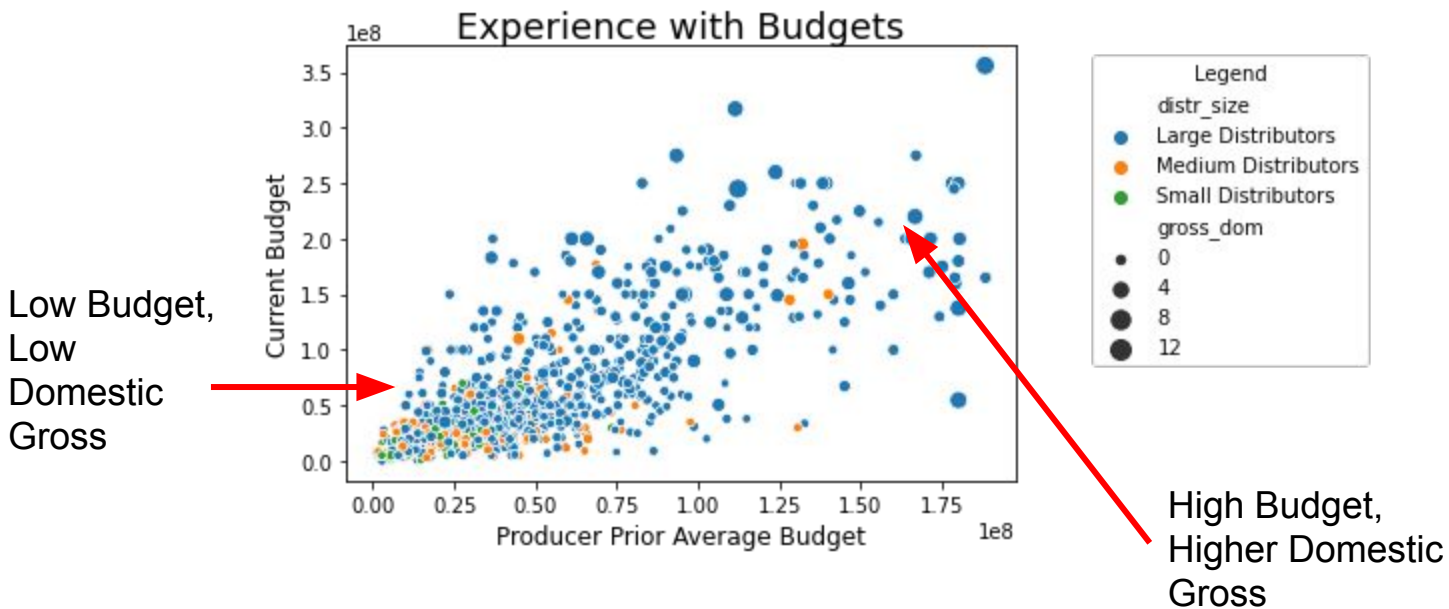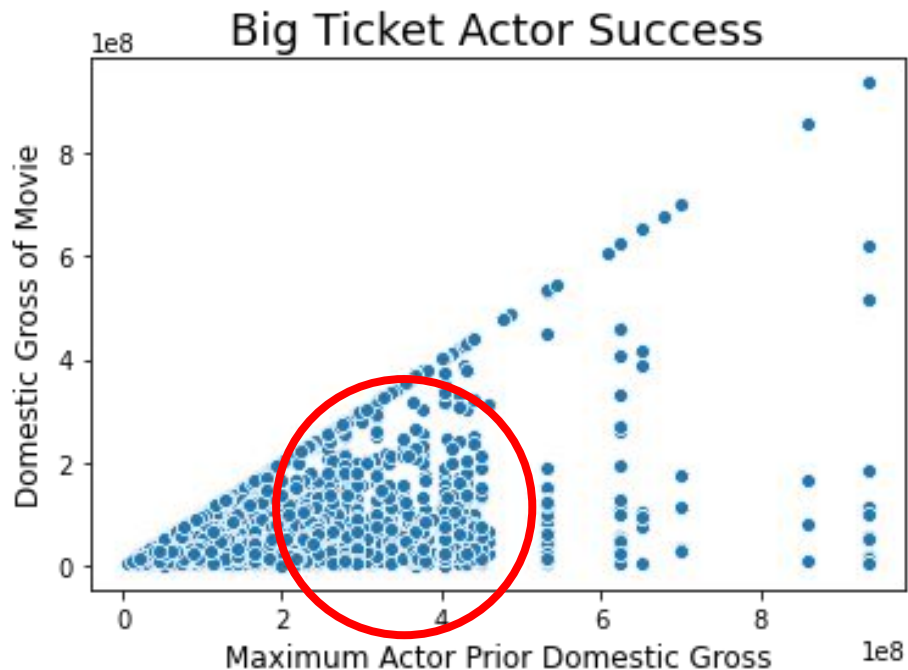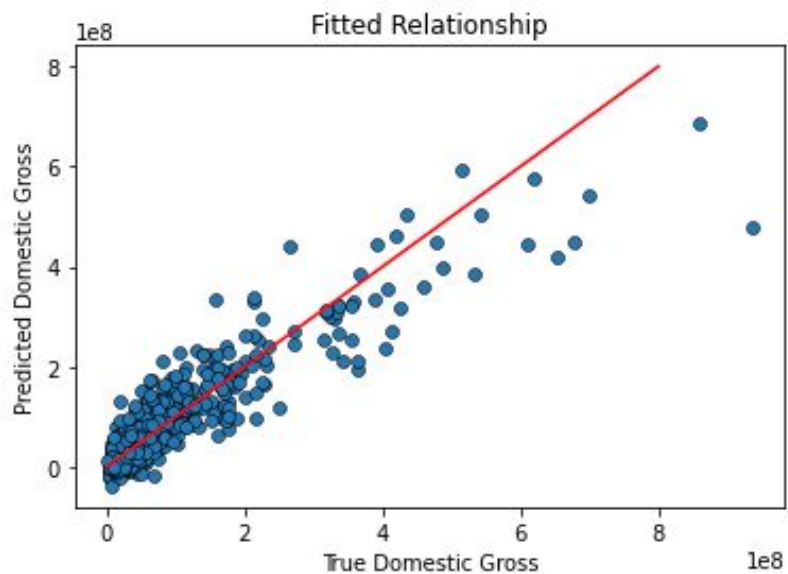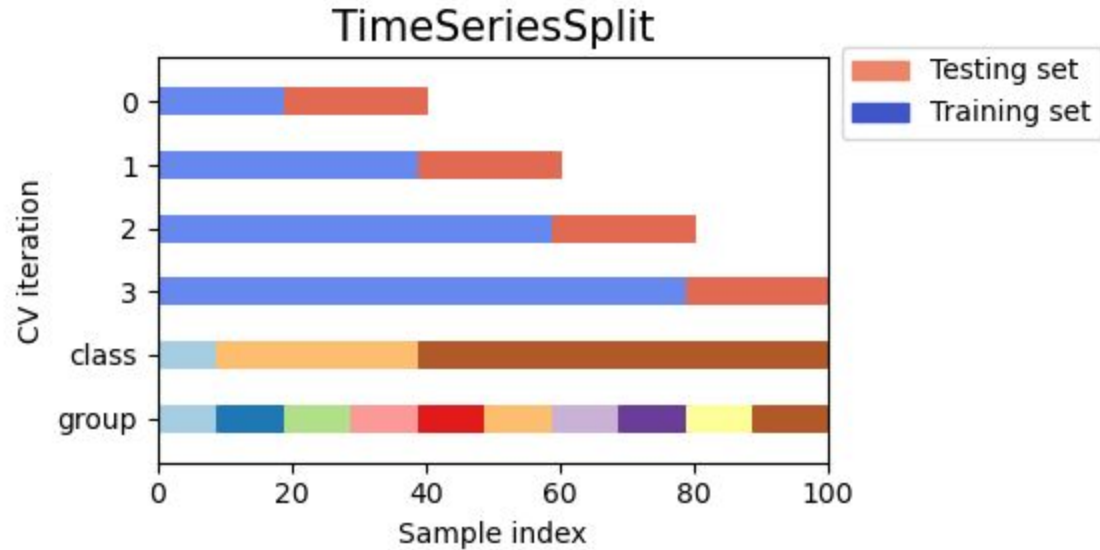
Does Money Make Money?

Maybe!

Results

Future Steps

❖ More data (Actual Actor salaries)
❖ Group Categorical Features like Genre
❖ Include Sequel or Franchise information

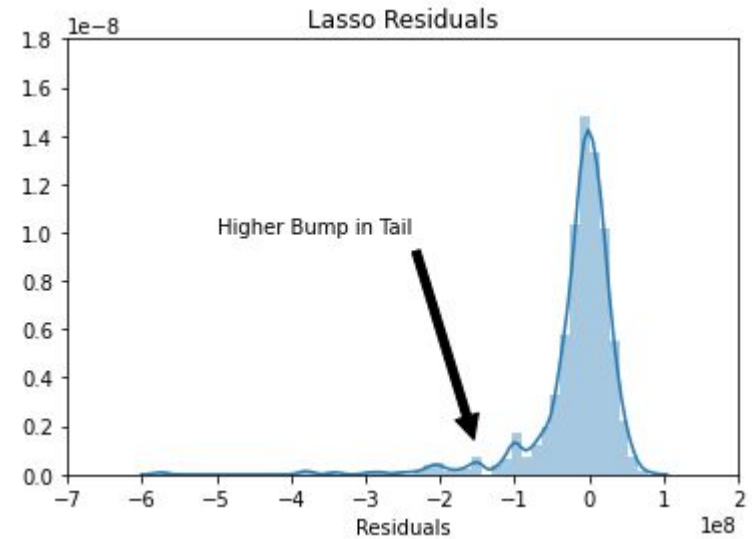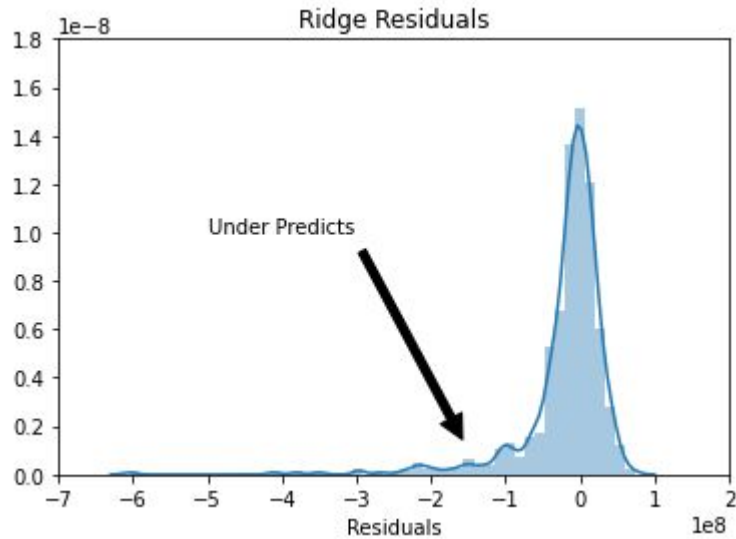# Questions?

Appendix - Cross Validation

Source: https://scikit-learn.org/

Appendix- Why not Lasso or Ridge?

Appendix- Metrics

	Linear Regression	RidgeCV	LassoCV
MAE:	31757101.609	33590271.259	33078139.356
RMSE:	48480221.371	63349602.927	61157570.744
R^2	0.79	0.69	0.71